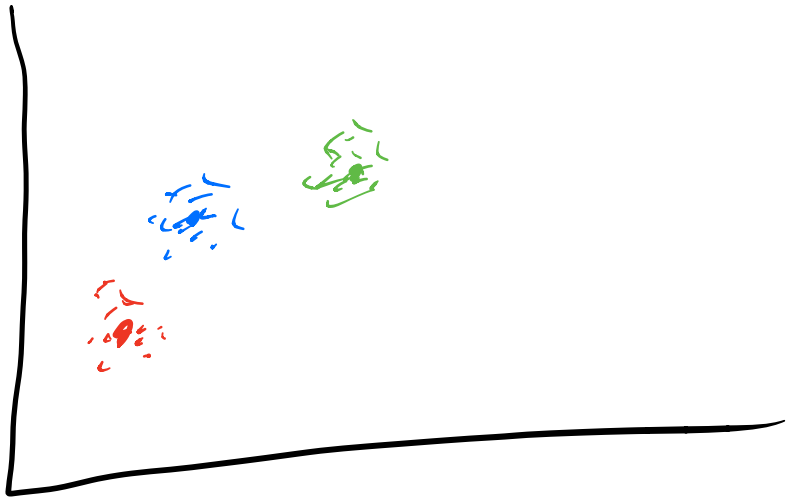
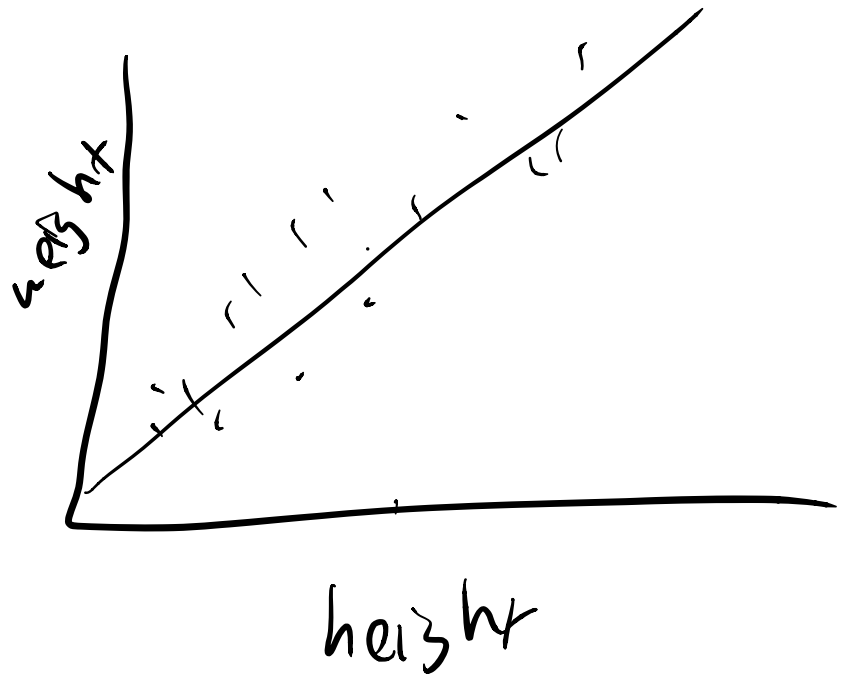
Regression

Ecological correlation means we expect greater correlation in absolute value among averages of groups than among ind. data points.



We use regression to take some X and predict y. We use our existing (x, y) data to "learn" how to do this through mean, sd, correlation



Procedure

1. Get all summary statistics for x, y (mean, sdx, sdy, corr).

2. Given a new x

3. Convert x to su

$$x_{su} = \frac{x - avg(x)}{sdx}$$

4. Multiply x_{su} by correlation

$$s = x_{su} \cdot r$$

5. Convert out of su into y

$$s \cdot sdy + avgy$$

All together

$$y_{pred} = \left(\frac{x_{new} - avg\,x}{sdx}\right) \cdot r \cdot sdy + avg\,y$$

Linear assumption

1) Linear, no outliers

2) x, y normal

3) homoscedasticity

Regression effect

we predict close to mean